John von Neumann Institute for Computing

Classification of Kinases: A Fast, Automated Structure-Based Approach

J. Lenz, Th. Margraf, Th. Lemcke, A. Torda

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 293-296, 2008.

http://www.fz-juelich.de/nic-series/volume40

Classification of Kinases: A Fast, Automated Structure-Based Approach

Jörn Lenz[1,2], **Thomas Margraf**[1], **Thomas Lemcke**[2], **and Andrew Torda**[1]

[1] Center for Bioinformatics, University of Hamburg, Bundesstrasse 43, 20146 Hamburg, Germany
E-mail: {jlenz, margraf, torda}@zbh.uni-hamburg.de

[2] School of Pharmacy, University of Hamburg, Bundesstrasse 45, 20146 Hamburg, Germany
E-mail: lemcke@chemie.uni-hamburg.de

We have been developing phylogenetic methods based on protein structures rather than sequence and fast enough to be applied to large families. A good example of a large family are kinases. Often, one believes in evolutionary relationships based on protein function, but one cannot see the relationships because the sequences have diverged so far. This is certainly the case with protein kinases. They are found in most forms of life, but with a tremendous spread of sequences and even differences in function. They are an ideal candidate for our approach since there are hundreds of known structures. Phylogenetic trees can be built automatically and they even map rather well to the biochemical annotations which were determined manually.

1 Introduction

Historically, phylogeny of large protein families has been based on sequence information. We have been developing methods, which are based on protein structure, but are still fast enough to be applied to large numbers of proteins. Here we consider the example of kinases, which are able to alter the activity of enzymes or other molecules by covalently attaching phosphate groups. This strategy usually denotes a response to chemical signals with some persistence, depending on reversibility and degradation mechanisms.

Kinases are central components in signal transduction networks and can be found in nearly all regulatory and metabolic processes in eukaryotes[1] and also many prokaryotes[2]. They play a major role in cell growth, division and controlled cell death, as well as in hormone response. Changes in human kinase activity can cause erroneous phosphorylation and trigger severe ailments such as cancer, diabetes or neurodegeneration[3]. Thus, kinases are suitable targets for the treatment of such diseases[4].

Understanding their evolution could help to explain the specific functions of individual kinases. It could assist in decoding signalling events and the emergence of pathologic biochemical processes. This might contribute to a more detailed insight of drug selectivity and drug cross reactivity and thus to the development of more effective drugs. It may also aid the selection of kinases used for drug screenings.

There is almost no significant sequence similarity between the more distant kinases, so it is difficult to build reliable sequence alignments. This superfamily, however, has been popular amongst crystallographers, so there is a wealth of solved structures. This makes it an ideal candidate for a structure-based phylogeny.

The only similar project in the literature was based on 31 kinase structures and required human intervention to construct a phylogenetic tree[5]. Here, we show how one can use many hundreds of structures to build a phylogeny completely automatically.

2 Methods

The list of kinase structures was assembled from successive structure searches[6]. Subsequently, a multiple alignment of those structures was computed using HANSWURST[7]. The guide trees, superimposed structures, and derived sequence alignments are analysed below. A more detailed description of the methods used is given by Margraf[7] and Lenz[8].

3 Results and Discussion

In this section we present an excerpt of the results of the Neighbour Joining clustering on RMSD values of pairwise superpositions (fig. 1). The multiple structure alignment and the implied multiple sequence alignment of the CMGC members show that the conserved features of kinases are appropriately superimposed. The most significant deviation from functional classification relates to the AGC kinases and the TK group. Firstly, 1h1wA is not clustered with any other member of the AGC group. Secondly, the G-protein coupled receptor kinase 1omwA is clustered with 1muoA from the "other" group. The superposition reveals that both structures share important features such as α-helix B[5] and could be structurally aligned. Thirdly, the AGC kinases 1cdkA and 1o6lA are clustered with 1phkA. The structural superpositions indicate that none of the alignments are unreasonable. Concerning the TK group members, only three of five were grouped together. However, they are interspersed with a TKL group kinase (fig. 1). The remaining two kinases of this group are clustered, but their distance to the CAMK group is closer than the distance to the other TK members. Another result of this work is the unusual placement of the TGFβR1 kinase 1b6cB. HANSWURST classifies this kinase as most closely related to 1m14A. Again, the structural superposition explains the result. The two structures are very similar. Additionally, HANSWURST clusters 1kwpA and 1csnA in one branch. Both kinases share most of the conserved structural features. Nevertheless, the structural superposition and the corresponding alignment of the sequences indicate that we aligned the proteins in a suboptimal manner. The p21-activated kinase 1 (PAK1) 1f3mC is placed close to the cluster of 1omwA and 1muoA. The second "other" kinase 1o6yA appears to be clustered with CAMK group kinases. A detailed analysis of alignments is given by Lenz[8].

To summarise, the tree is more than reasonable without any serious misalignments. Compared to the previous literature attempt[5] at classification, there are some differences, especially with the AGC family, but they appear justified. We should only agree with the literature classification of the CMGC and AK groups. Furthermore, our methodology handled an order of magnitude more structures and was fully automatic.

4 Conclusion

This work concentrated on kinases, since one can compare against literature classifications and biochemical data. Because it is fully automatic and scales well, it can be expanded to even larger families. This means we are now considering even more distantly related proteins and testing the approach on other large protein families.

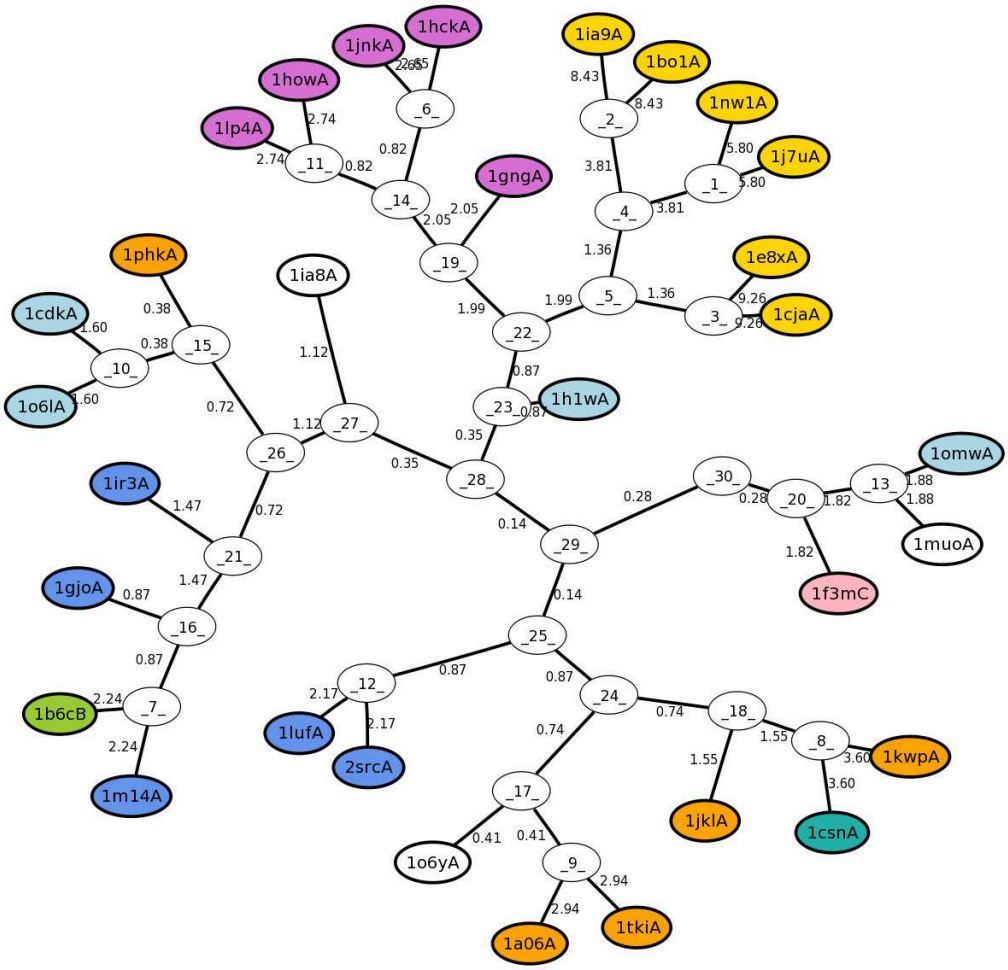

Figure 1. *HANSWURST proposed phylogeny for the literature list of recognized kinases; edges are labeled with estimated RMSD values according to clustering method; edges are not drawn to scale; atypical kinases (gold), typical kinases (cornflowerblue), AGC kinases(lightblue), tyrosine kinase like kinases (yellowgreen) depicting the most diverse group, CK1 kinases (seagreen), STE (lightpink), CMGC (magenta), CAMK (orange), uncoloured leaves depict members of the "other" group.*

References

1. S Hanks, *Genomic analysis of the eukaryotic protein kinase superfamily: a perspective.*, Genome Biol, **4**, no. 5, 111, 2003.
2. P Kennelly, *Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry.*, Biochem J, **370**, no. Pt 2, 373–389, 2003.
3. J Goldberg, G Manning, A Liu, P Fey, K Pilcher, Y Xu, and J Smith, *The dictyostelium kinome–analysis of the protein kinases from a simple model organism.*, PLoS Genet, **2**, no. 3, e38, 2006.

4. M Noble, J Endicott, and L Johnson, *Protein kinase inhibitors: insights into drug design from structure.*, Science, **303**, no. 5665, 1800–1805, 2004.
5. E Scheeff and P Bourne, *Structural evolution of the protein kinase-like superfamily.*, PLoS Comput Biol, **1**, no. 5, e49, 2005.
6. G Schenk, T Margraf, and AE Torda, *Protein sequence and structure alignments within one framework.*, Algorithms Mol Biol, **3**, no. 1, 4, 2008.
7. T Margraf, "Fast multiple alignment of protein structures", Diploma thesis, Center for Bionformatics, University of Hamburg, June 2007.
8. J Lenz, "Fast and efficient structure-based classification of kinases", Diploma thesis, Center for Bionformatics, University of Hamburg, March 2008.